Intrinsic Fluorescence-Based *at Situ* Soft Sensor for Monitoring Monoclonal Antibody Aggregation

Kaveh Ohadi, Raymond L. Legge, and Hector M. Budman
Dept. of Chemical Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada

DOI 10.1002/btpr.2140
Published online July 22, 2015 in Wiley Online Library (wileyonlinelibrary.com)

Intrinsic fluorescence spectroscopy, in conjunction with partial least squares regression (PLSR), was investigated as a potential technique for online quality control and quantitative monitoring of Immunoglobulin G (IgG) aggregation that occurs following exposure to conditions that emulate those that can occur during protein downstream processing. Initially, the impact of three stress factors (temperature, pH, and protein concentration) on the degree of aggregation determined using size exclusion chromatography data, was investigated by performing a central composite designexperiment and applying a fitting response surface model. This investigation identified the influence of the factors as well as the operating regions with minimum propensity to induce protein aggregation. Spectral changes pertinent to the stressed samples were also investigated and found to corroborate the high sensitivity of the intrinsic fluorescence to conformational changes of the proteins under study. Ultimately, partial least squares regression was implemented to formulate two fluorescence-based soft sensors for quality control—product classification—and quantitative monitoring—concentration of monomer. The resulting regression models exhibited accurate prediction ability and good potential for in situ monitoring of monoclonal antibody downstream purification processes.
© 2015 American Institute of Chemical Engineers *Biotechnol. Prog.*, 31:1423–1432, 2015
Keywords: immunoglobulin G aggregation, bioprocess monitoring, fluorescence spectroscopy, partial least squares, soft sensor

Introduction

Monoclonal antibodies (MAbs) are the dominant products in the biologics market[1] with the immunoglobulin Gs (IgGs) outpacing other bioproducts.[2] Hydrophobic recombinant proteins (e.g., IgGs) are inherently susceptible to aggregation[2,3] during downstream processing following exposure to stress factors such as pH, temperature, and agitation.[4,5] Protein aggregation is considered as the most common hindrance for process development[6] with such adverse effects as loss of efficacy and/or provoking an immunogenic response.[4,5,7,8] These issues are fueling the development of *in situ* techniques that enable real-time accurate quality and quantity control of protein aggregation.

Conventional techniques for monitoring protein aggregation are comparatively time consuming, while intrinsic fluorescence spectroscopy is a fast, noninvasive, and nondestructive technique with high sensitivity and signal to noise ratio, which is amenable to online monitoring.[7,9–11] Fluorescence spectroscopy has been identified as a plausible technique for monitoring conformational changes and characterization of protein tertiary structure.[5,9,12]

Three intrinsically fluorescent aromatic amino acids (phenylalanine, tyrosine, and tryptophan), in the protein chain, have been found to be sensitive to the micro-environment they are exposed to. During protein disruption and aggregate formation, the surrounding environment of the fluorophores change. Accordingly, changes in their fluorescence behavior can be exploited to draw inferences on conformational changes as well as native and non-native characteristics of the protein structure.[5,10,12] The higher quantum yield and extinction coefficient of tryptophan makes it a desirable probe to track protein aggregation.[9,13] The emission maxima and intensity of the tryptophan shift is sensitive to its environment. This shift indicates whether tryptophan is exposed to the surrounding solvent or buried within the protein.[9,13] Multiwavelength intrinsic fluorescence has previously been investigated for use as a soft sensor for monitoring α-lactoglobulin and β-lactoglobulin solubility under stressed conditions[10,14] and for discriminating between different types of cheeses.[15] Kumar *et al.* (2005) demonstrated the suitability of second-derivative fluorescence spectra of tryptophan to identify subtle structural changes in β-lactoglobulin and interferon alpha-2a upon exposure to various solvent conditions. Second-derivative fluorescence has also been implemented to qualitatively assess MAb conformational changes under thermal, pH, and solvent stressed conditions.[9] However, the utilization of intrinsic multi-wavelength fluorescence spectroscopy in the development of a soft sensor for both qualitative and quantitative monitoring of MAb aggregation has not been explored.

The implementation of chemometric methods has been explored in the context of fluorescence-based soft sensor development.[11,16,17] Partial least squares regression (PLSR), one of the most commonly used data exploratory techniques,[18]

Correspondence concerning this article should be addressed to Hector M. Budman at hbudman@uwaterloo.ca.

© 2015 American Institute of Chemical Engineers

was utilized to draw indirect inferences from the fluorescence spectra to generate predictive models for monitoring aggregation.

The objective of this study was to develop soft sensors based on *intrinsic* fluorescence, coupled with PLSR, for *at-line* quality control—classification of product—and also for quantitative monitoring—prediction of monomer concentration—at different stages of a downstream purification process. For simplicity and to have better control of experimental conditions, the measurements were collected by exposing the samples to operating conditions analogous to those that occur during protein purification. Given the diverse aggregation patterns induced by various stress factors imposed along downstream processes, IgG samples of different concentrations were subjected to different temperature and pH conditions. Temperature and pH changes are commonly used to achieve separation of proteins.[19] A response surface method (RSM) was formulated to describe the aggregation patterns induced by different stress factors. RSM was performed to better comprehend the effect and interaction of the stress factors as well as to determine the regions with minimum propensity to cause aggregation. Additional experiments were also conducted to produce a more diverse data set for soft sensor development. High-pressure size exclusion chromatography (HP–SEC) was utilized for IgG-sample fractionation and independent quantification. Finally, PLSR-based models were formulated between HP-SEC measurements and fluorescence spectra collected for the corresponding samples. These models can be used as a soft sensor for predicting monomer concentration and inferring conformational changes from measured fluorescence spectra.

Materials and Methods

Sample preparation

ChromPure Human IgG (Cedarlane, ON) with a concentration of ~11.3 mg/mL in 0.01*M* phosphate buffer saline (PBS) stored at pH 7.6 was utilized for the experiments. To study the impact of temperature, pH, and protein concentration on the aggregation, an asymmetric central composite design (CCD) with 4 center points was implemented (Table 1). Because of the sample limitations and a need for a certain combination of conditions for soft sensor development purposes, it was not possible to perform symmetric CCD. Sample dilution was performed using HyClone 0.01*M* PBS (Fisher Scientific, ON) at pH 7. For the experiments performed at different pH levels, the pH of the buffer was adjusted using HCl (1*M*) and NaOH (1*M*) prior to sample dilution. Samples were thermally stressed at different temperatures in a water bath for 20, 40, and 60 min then cooled for 20 min at room temperature in a water bath before being centrifuged at 400*g* for 5 min. The resulting supernatant was decanted and used for the fluorescence and HP-SEC measurements. For the purpose of developing a soft sensor for *in situ* monitoring of protein aggregation, additional experiments were carried out (Table 1).

Table 1. Summary of Experimental Design

Experiment No.	Temperature (°C)	pH	Concentration (mg/mL)
1	70	5.1	0.2
2	70	8.1	0.2
3	70	8.1	1
4	70	5.1	1
5	50	5.1	0.2
6	50	8.1	0.2
7	50	8.1	1
8	50	5.1	1
9	80	7	0.5
10	40	7	0.5
11	60	9.1	0.5
12	60	3.1	0.5
13	60	7	1.5
14	60	7	0.1
15-18	60	7	0.5
19	60	7	1
20	60	7	0.2
21	60	8.1	0.5
22	60	5.1	0.5
23	60	5.1	0.2
24	60	5.1	1
25	60	3.1	0.2
26	60	3.1	1
27	60	9.1	0.2
28	70	7	0.5
29	70	7	0.2

Experiment 1–18: Asymmetric central composite design (CCD) with four center points and experiment 19–29: Complimentary set of experiments.

High-pressure size exclusion chromatography

The resulting supernatant was analyzed by HP-SEC on an Agilent 1200 chromatography system (Palo Alto, CA) equipped with a UV detector. Two-hundred microliter of sample were injected into a Speax Zenix-C SEC-300 (Sepax Technologies, Newark, DE) column with a flow rate of 1 mL/min for 20 min with a mobile phase of 1.5*M* PBS at pH 7.4 and detected at 280 nm. Measurements were carried out in duplicate. A calibration curve was prepared with various concentrations of pure IgG_1 ranging from 1.5 to 0.01 mg/mL. Protein concentration calculations were based on the area under the curve (AUC) pertinent to the peak of the monomeric form of the IgG_1. To avoid errors related to elution caused by column saturation or attachment of highly hydrophobic molecules, a new calibration curve was generated after column regeneration on a regular basis. The peaks were attributed to different oligomeric structures based on the relative residence time as compared to the information reported in the Ref.[3]. The aggregation percentage was defined as the difference between the concentration of the protein in monomeric form before and after being subjected to the imposed stresses. This difference was then normalized with respect to the initial monomer concentration of unstressed sample. As such, the calculated aggregation percentage is equivalent to a measure of product loss.

Fluorescence spectroscopy

Multiwavelength fluorescence spectra of the supernatant were acquired at room temperature utilizing a Cary Eclipse spectrofluorometer (Palo Alto, CA) equipped with a Peltier multicell holder in 700 μl far UV quartz cells with path length of 10 mm (Mandel Scientific, ON). The measurements were collected over an excitation range from 260 to 350 nm at 5 nm increments and emission range from 280 to 450 nm with 1 nm increments. The photomultiplier tube (PMT) voltage of 600 V, slit width (SW) for excitation and emission of 5 nm, and scanning rate of 600 nm/min were set for signal acquisition. The measured fluorescence spectra was arranged in a form of a matrix (19 excitation × 121 emission) known in the literature as an excitation emission matrix (EEM).

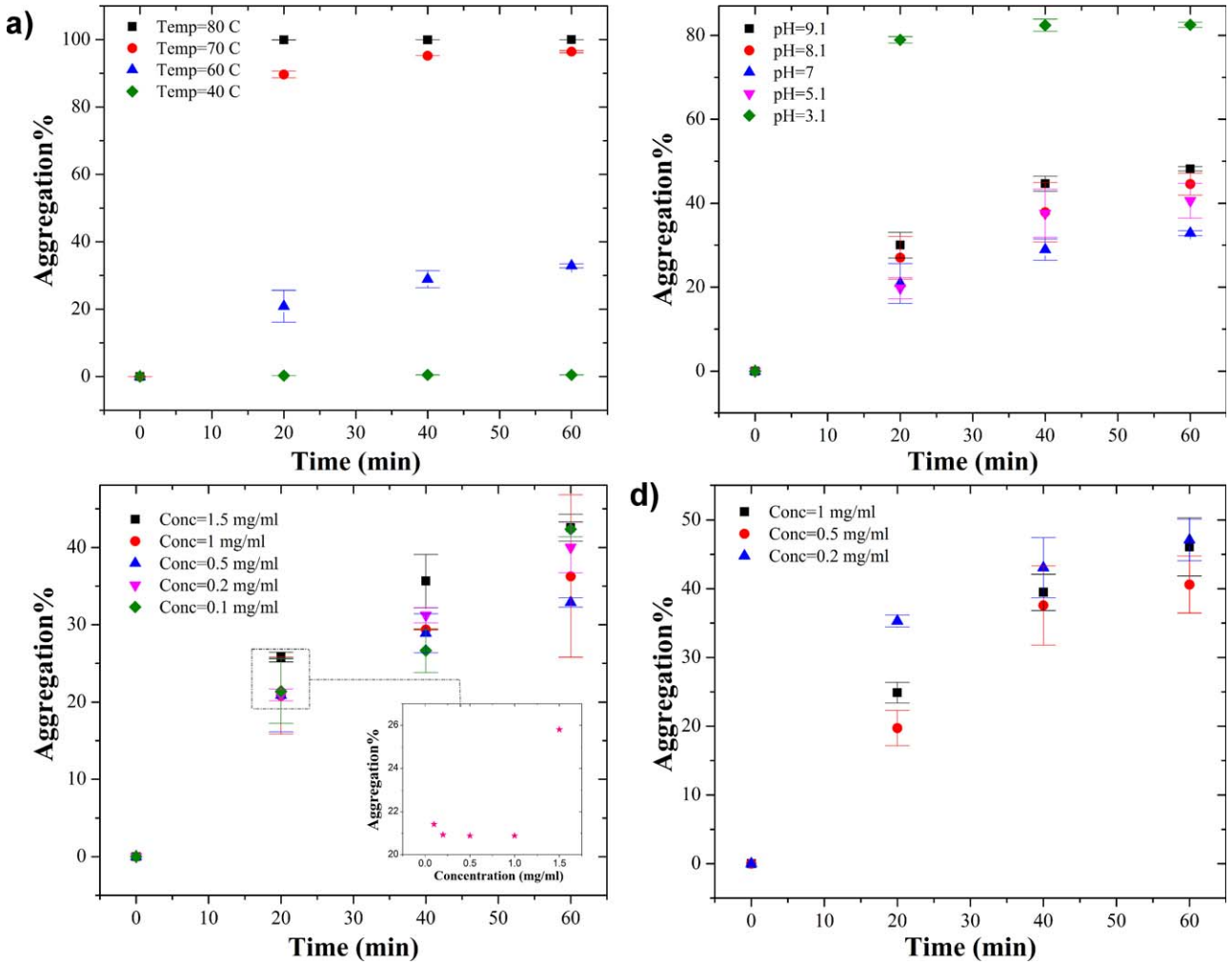

Figure 1. Temporal evolvement of the degree of aggregation under stressed conditions: (a) For samples with 0.5 mg/mL initial IgG at pH = 7 under different temperatures; (b) For samples with 0.5 mg/mL initial IgG stressed at 60°C under different pHs; (c) For samples with different initial IgG concentrations stressed at 60°C and pH = 7; (d) For samples with different initial IgG concentrations stressed at 60°C and pH = 5.1.

Since large insoluble aggregates were eliminated by centrifugation and the concentration of Mab was low, fluorescence quenching was not a concern.

Chemometric analysis

With the goal of developing empirical models (soft sensors) based on multiwavelength fluorescence spectra for quality control and quantitative monitoring, the PLSR[18,20] technique was applied. The input data matrix (X) is formed by individual re-arrangement of EEMs of samples into row vectors followed by their row-wise attachment. Sample duplicates were included in the X matrix separately. The formation of the response matrices (Ys) and the preprocessing methods for each soft sensor is explained separately in the corresponding sections. The optimal number of latent variables (LVs) is obtained by minimizing the root mean square error of cross validation (RMSE-CV) to avoid model overfitting. In the current study, a random subset routine with 20 iterations was carried for cross-validation (CV). Chemometric analyses were performed utilizing the PLS-Toolbox 7.0.3 (Eigenvector Research, Manson, WA) running in the MATLAB 8.0.0 (Mathworks, Natick, MA) platform.

Results and Discussion

Impact of stress factors

To better understand the impact of stress factors and investigate diverse aggregation patterns, an asymmetric central composite design with a complementary set of experiments (to encompass a wide range of stress factors) was conducted. Additionally, the relevance of the tryptophan fluorescence-spectra of the samples for tracking Mab conformational changes was thoroughly studied. It is worth noting that the samples that were not subjected to stress factors contained less than 5% dimers, which are neglected in the current study.

Temperature plays a crucial role to evoke aggregation. Higher temperatures contribute to the protein conformational changes, exposure of hydrophobic regions of the polypeptide chain to the external medium, and consequently to the formation of aggregates.[4,5]

Figure 1a exhibits the changes in aggregation percentage of IgG samples (pH = 7 and concentration = 0.5 mg/mL) with time at different temperatures. From the results it can be observed that the propensity for aggregation is proportional to the temperature and occurs when samples are heated above a minimum temperature. For instance, over

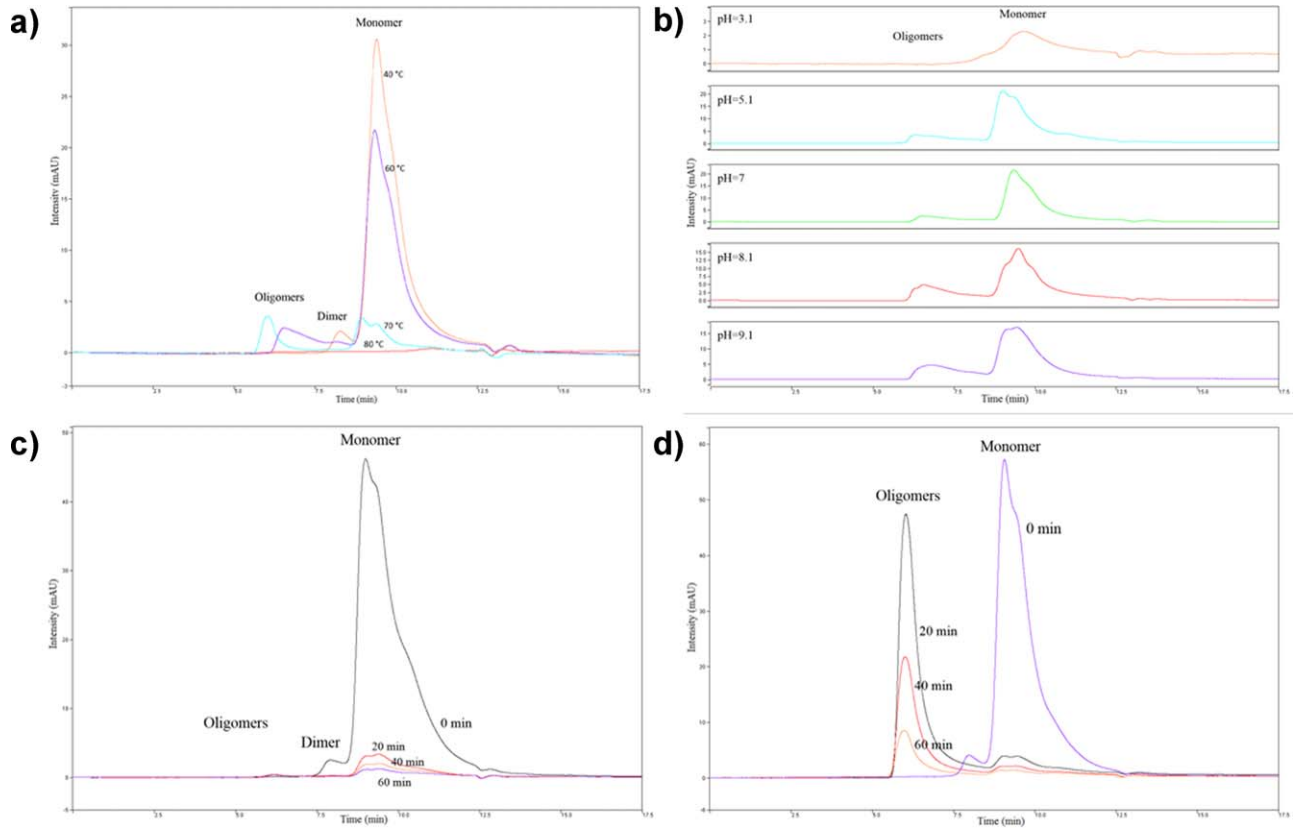

Figure 2. Chromatograms of samples obtained from SEC: (a) For samples with 0.5 mg/mL initial IgG at pH = 7 after 20 min of treatment under different temperatures; (b) For samples with 0.5 mg/mL initial IgG stressed at 60°C for 20 min under different pHs; (c) For samples with 1 mg/mL initial IgG and pH = 5.1 stressed at 70°C at different sampling time; (d) For samples with 1 mg/mL initial IgG and pH = 8.1 stressed at 70°C at different sampling time.

20% of the IgG monomer is aggregated after 20 min of heat treatment at 60°C, while at 40°C the level of aggregation is negligible (chromatogram for the unstressed sample was identical to that of the sample exposed to 40°C). Higher temperatures (70 and 80°C) resulted in a significant loss of the monomer structure as well as induced formation of large insoluble oligomers that precipitate as supported by the chromatograms in Figure 2a for samples subjected to these higher temperatures (pH = 7 and concentration = 0.5 mg/mL) for 20 min. This phenomenon is attributed to the formation of greater levels of unfolded IgG at higher temperatures.[7] Regardless of the initial IgG concentration and pH, similar trends were observed for temperature-induced aggregation (results not shown), which is in accordance with the Ref. 5,7. The typical melting point of IgG is reported to be above 70°C supporting these findings.[21]

The impact of pH on a protein is dependent on the amino acid composition as constituent amino acids become protonated or deprotonated. This changes the conformational stability and the free energy of unfolding.[5,6] On the basis of the experimental results obtained at different pHs, it was concluded that samples with a pH close to neutral levels had a very small tendency to aggregate while deviations from pH 7 resulted in an increase in the rate of aggregation (Figure 1b). Figure 2b is a comparison between the heat-treated samples, at 60°C, 0.5 mg/mL after 20 min, at different pHs. It can be seen that under acidic conditions (pH = 3.1) the rate of loss of monomer is significantly higher. The substantial increase in the intensity of tryptophan fluorescence is in agreement with this conclusion. Additionally, it was observed that at 70°C and an initial IgG concentration of 1 mg/mL, the sample with pH 5.1 experienced more pronounced loss of monomer and formation of insoluble oligomeric structures over time, in contrast with samples subjected to pH 8.1 that exhibited a higher tendency of forming trimer and tetramers (Figures 2c,d).

Although protein aggregation is generally found to increase with protein concentration,[4,6] the pattern of protein aggregation is found to be inconsistent (for the concentration range of this study) under different pH conditions. For instance, at pH 7 a sample with concentration of 1.5 mg/mL experienced the greatest loss of monomers and at the same time formed more trimers/tetramers after 20 and 40 min exposures to 60°C (Figure 1c), while at pH 5.1 the aggregation percentage of the sample with initial concentration of 0.2 mg/mL outpaced that of 0.5 and 1 mg/mL samples (Figure 1d). In spite of the high level of aggregation observed under acidic conditions (pH = 3.1), the sample with an initial concentration of 1 mg/mL (at 60°C after 20 min) aggregated 10% and 7% more than samples with initial IgG of 0.5 and 0.2 mg/mL respectively (results not shown).

Figure 3a provides the tryptophan fluorescence (Excitation = 280 nm) of IgG samples (pH = 7 concentration = 0.5 mg/mL) stressed at 60°C over a 60 min time span. An increase in the tryptophan intensity over time and a red-shift of peak maxima are indicative of protein unfolding and the formation of oligomeric structures that are observed in the HP-SEC results for the same samples during the course of heat induced

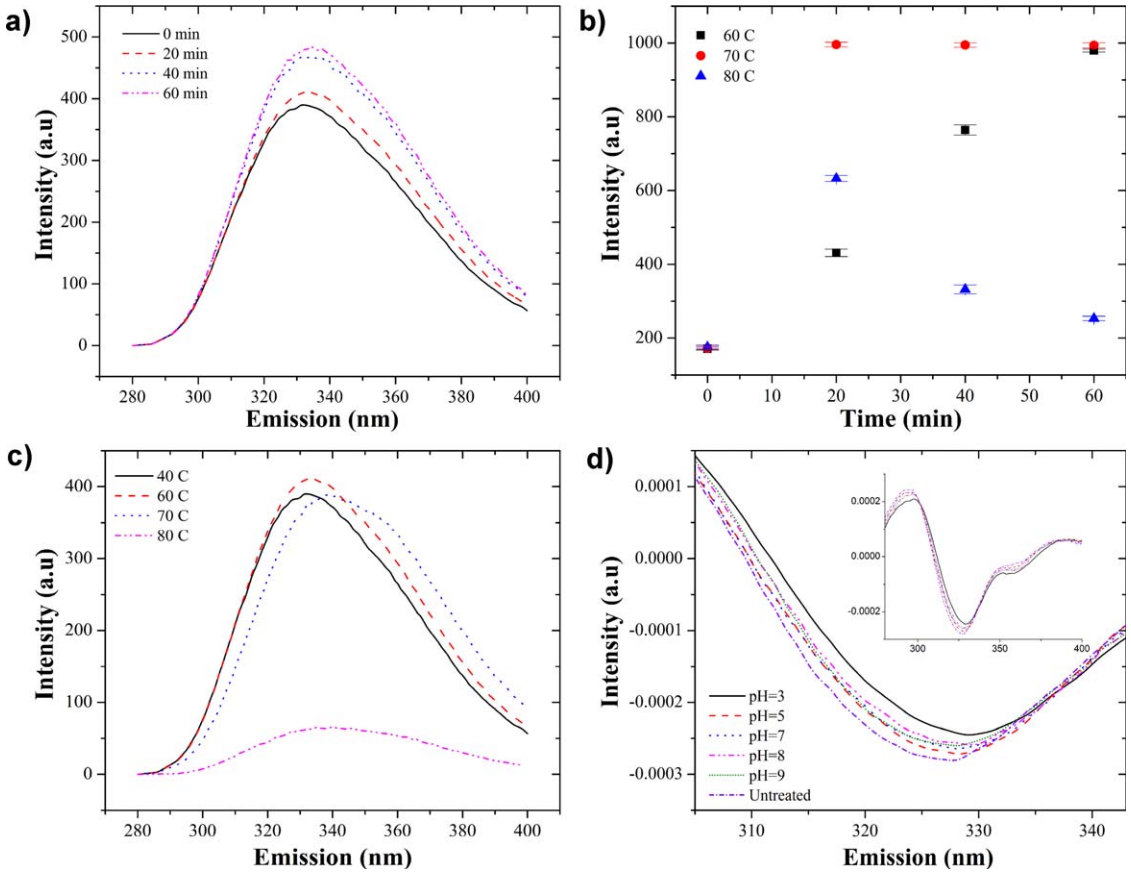

Figure 3. Fluorescence signal of stressed samples: (a) Emission spectra (at excitation 280 nm) for samples with 0.5 mg/mL initial IgG and pH = 7 stressed at 60°C at different sampling time; (b) Intensity of scattering-maxima (at excitation 305 nm) for samples with 0.5 mg/mL initial IgG and pH = 7 stressed at different temperatures over time; (c) Emission spectra (at excitation 280 nm) for samples with 0.5 mg/mL initial IgG and pH = 7 stressed after 20 min at different temperatures; (d) Second derivative of emission spectra (at excitation 280 nm) for samples with 0.5 mg/mL initial IgG after 20 min treatment at 60°C at different pHs.

aggregation. A pronounced change in the tryptophan fluorescence between 20 and 40 min and a subtle change occurring between 40 and 60 min are in agreement with the results in Figure 1a. Rayleigh light scattering is an incident light, occurring at equal excitation and emission wavelengths and it is generally attributed to the presence of suspended particles and their size.[11] The scattering intensity (Excitation = Emission = 305 nm) for the aforementioned samples continually increases (Figure 3b) over time and is attributed to the formation of larger oligomers in the samples in agreement with the protein fractionation results (results not shown). The rationale behind selecting the scattering at an Excitation = Emission = 305 nm is that at lower wavelengths, because of high absorbance of energy by tryptophan, a low signal to noise ratio was observed for scattering. At higher wavelengths, a lack of energy absorption by the fluorophores leads to higher energy of scattering that saturates the detector thus causing a loss of sensitivity. At higher temperatures (80°C) the overall protein concentration decreases because of the precipitation of large oligomers. This is evident in the sudden reduction in the tryptophan fluorescence intensity and a higher red-shift in the peak maxima of samples (pH = 7 concentration = 0.5 mg/mL) after 20 min of heat treatment (Figure 3c) as well as the scattering levels (Figure 3b), which agrees with the HP-SEC results in Figure 2a. Abbas *et al.*[9] argued that the second derivative of the normalized tryptophan signal can be implemented to track induced conformational changes. Figure 3d provides the second derivative (calculated using the Savitzky–Golay algorithm) of the normalized emission signal obtained at an excitation of 280 nm for a sample with 0.5 mg/mL initial IgG after 20 min of heat treatment at 60°C under various pHs. A reduction in value and a shift to higher wavelengths for the peak (minima) at ~327 nm is an indicator of the loss of the native structure of IgG with larger disruption occurring for the samples exposed to pH = 3. From Figure 3d a similarity can be found between the aggregation pattern of samples at pH = 8 and 9 and pH = 5 and 7, as their minima approach each other, which is in agreement with previous observations inferred from Figures 1b and 2b and agrees with the findings of Abbas *et al.*[9] The subtle changes in the intrinsic fluorescence-spectra of stressed samples reinforce the capability of fluorescence spectroscopy to serve as a tool for quantitative and qualitative monitoring of aggregation.

Response surface model

An RSM was developed to describe the impact of stress factors on the percentage of aggregation. The experimental error is the total error calculated from measurements of monomer concentration with SEC for replicated samples of the center point and is estimated to be ~12%. The response variable was defined as the percentage of aggregation after 20 min. For better accuracy, a logarithmic transformation of the response variable was used for model calibration. Since

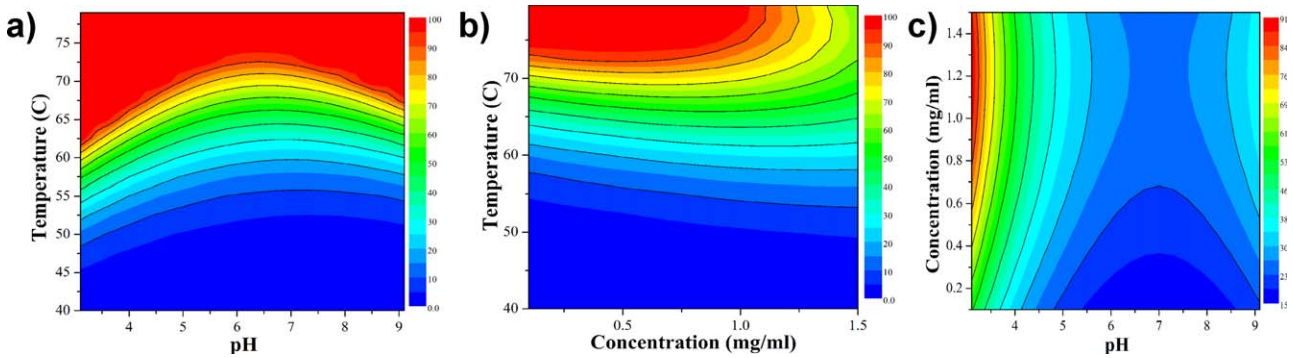

Figure 4. Contour plots of fitted surface model: (a) Temperature versus pH at initial IgG of 0.5 mg/mL; (b) Temperature versus concentration at pH = 7; (c) Concentration versus pH at 60°C.

Table 2. Analysis of Variance (ANOVA) of the Final Response Surface Model (RSM) After Excluding the Insignificant Factors

	Degree of Freedom (df)	Sum of Square (SS)	Mean Square (MS)	$F_{Observed}$	F_{Actual}
SSE	10	394.72	39.47		
SSLF	7	373.65	53.38	7.596	8.89
SSPE	3	21.08	7.02		

the aggregation was bounded to 0% and 100%, the model output was forced to be equal to 0 for prediction values below 0 and 100 for prediction values above 100. Analysis of variance (ANOVA) along with a *F*-test was performed for various model structures to evaluate the lack of fit for each proposed model. Consequently, a function containing quadratic, interactions, and linear terms was found to be the best fit to the experiments. ANOVA with *t*-test was then performed to determine significant effects and fine-tune the final model.

Table 2 summarizes the ANOVA of final RSM including only the significant effects. Temperature was found to have the most profound impact on aggregation. It was also noted that the impact of pH and concentration is slightly dependent on the temperature, while no considerable interaction was observed between pH and protein concentration. Additionally, a curvature effect was noted for all three effects which were evident from the conclusion drawn previously. Figure 4 provides the contour plots of aggregation percentage after 20 min of treatment, obtained from the surface model, with respect to two factors while the third is kept constant. These contour plots can be employed to evaluate the operating regions where the propensity of aggregation formation is lower. For example, it was found that the tendency of aggregation at pH values ranging from 6 to 7 is comparatively lower than pH levels above or below this range at any temperature (Figure 4a) and monomer concentration (Figure 4c) investigated. At a protein concentration of 0.5 mg/mL, the impact of basic pH levels on aggregation is more substantial at higher temperatures, as compared to lower ones (Figure 4a). A similar behavior was observed for other concentrations. Regardless of the sample pH, the rate of aggregation at different concentrations was found to be dependent on temperature (Figure 4b for pH = 7). To improve the extrapolation accuracy of the surface model at high temperatures and monomer concentrations, more data were acquired for further calibration.

Soft sensor for quantitative monitoring

An empirical model (soft sensor) was developed based on multi-wavelength fluorescence spectra and utilizing the PLSR method, in order to predict the concentration of the monomeric IgG of pure (unstressed) and stressed samples. This soft sensor can be used *at situ* to track desirable product concentrations as well as monitoring loss of monomers. Since the aim of this soft sensor is to estimate the monomeric content, only the part of spectra pertinent to the tryptophan peak was implemented as input to the model. Thus, the scattering region, which is believed to provide information about the suspended particles and is probably more relevant to oligomeric structures, was eliminated from the spectra using an in-house developed MATLAB program. The concentrations of the IgG in monomeric structure obtained via protein fractionation (as explained in section "High-pressure size exclusion chromatography") were appended together to form the output matrix (Y). To ensure the model-prediction accuracy experiments 2, 13, 17, 20, 23, and 25 (Table 1) were randomly selected and eliminated from the calibration set. These experiments were then used to test the prediction accuracy of the resulting model. Five LVs were found to minimize the RMSE-CV resulting in a minimal error of 0.053 mg/mL of monomer concentration, which is equivalent to ~16% of the average monomer-IgG concentration; in accordance with 13% SEC measurement error. The resulting model was capable of capturing roughly 100% of variation in the input (X) and simultaneously explains 97.1% of the output (Y) for the calibration set which is deemed highly reasonable considering the R^2_{CV} equal to 96.9%. The resulting model predicted the test set with a precision of R^2_P = 97.6% that verifies the accuracy of the cross-validation procedure necessary to avoid model over-fitting. The model precision is depicted in Figure 5a where the calibration and validation sets are aligned closely to a 45° line with the residuals being approximately normally distributed (Figure 5a inset). The first latent variable of the PLSR model accounts for 98% and 57% of variance of the input and output data sets, respectively. Although the second LV only contributes to 1% of variation in the X matrix, it can describe roughly 34% of the response matrix Y. Figure 5b shows the scores of LV2 versus LV1 with the 95% confidence region. As shown in Figure 5b, the scores on the LV1 are discriminated into four regions that correspond to a different range of monomer concentrations. This correlation becomes more apparent in Figure 5c that shows the monomer concentration versus LV1 scores. In spite of the group

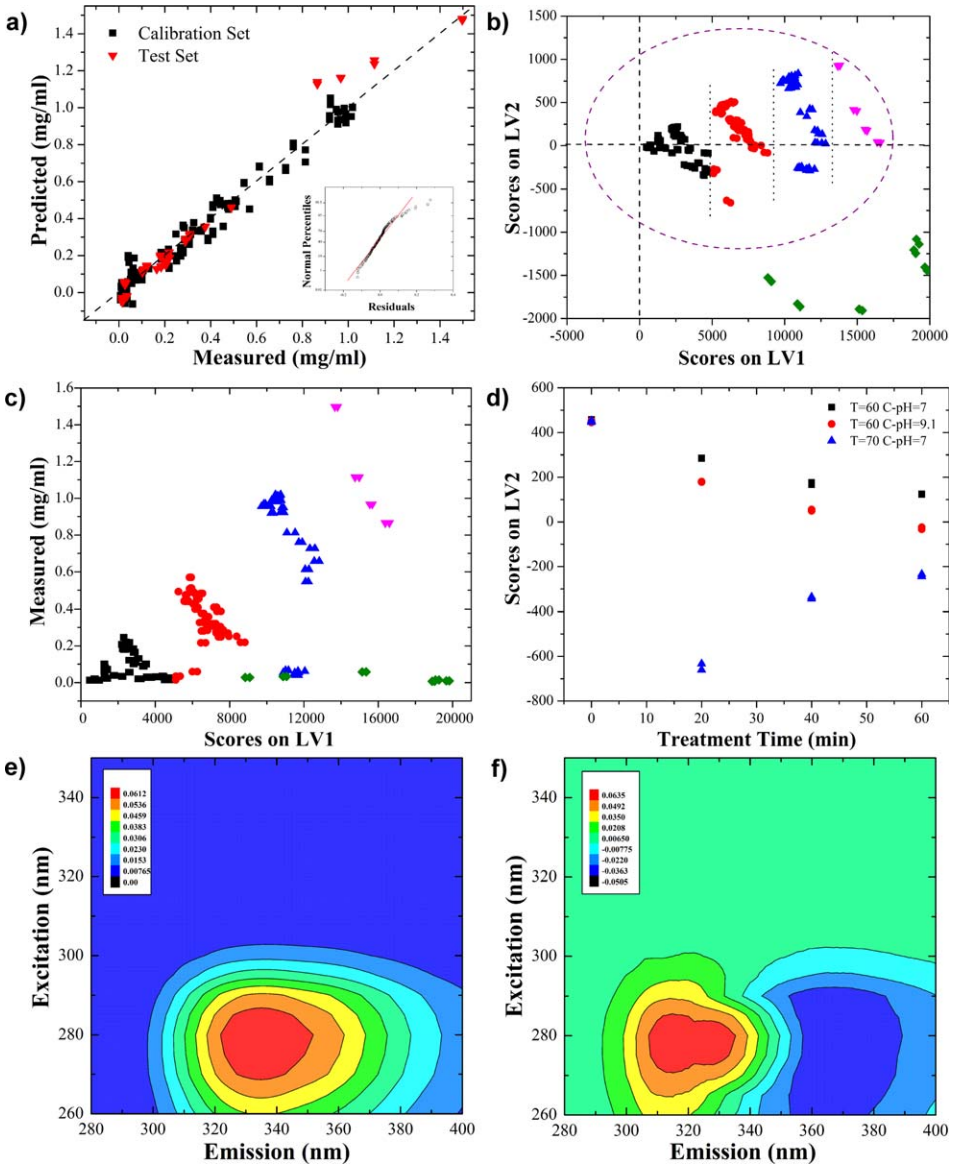

Figure 5. Results of the soft sensor for quantitative analysis: (a) Model predictions versus measured values for calibration and test set and normal percentiles of residuals (inset); (b) Scores of LV2 versus LV1; (c) Measured values versus scores on LV1; (d) Scores on LV2 for three different experiments over time; (e) Contour plot of the first loading; (f) Contour plot of second loading.

of outliers with extremely low LV2-scores (Figures 5b,c), a correlation is observed between monomer concentration and the scores on the LV1. The outliers observed in Figure 5b have very low leverage; hence, imposing a minimal effect on the model accuracy. Figure 5d shows the trajectories of LV2-scores for three different experiments with an initial IgG concentration of 0.5 mg/mL. For experiments performed at 60°C at pH 7 and 9.1, the LV2-scores are constantly decreasing, which is in agreement with the trend of formation of oligomers noted earlier in section "Impact of stress factors". On the other hand, for the experiment performed at 70°C and pH 7, an increase in LV2-scores is observed after reaching a minimum that can be explained by the formation of precipitates that were eliminated by the centrifugation step performed before the measurement.

The physical relevance of the latent variables can be explained by investigating the matrix of loadings. By performing an inverse of the matrix transformation explained in section "Chemometric analysis", the loadings pertinent to each latent variable were converted to the form of an EEM. Figures 5e,f are the contour plots of the first and second LVs. The first loading (Figure 5e) is correlated to the region corresponding to tryptophan with an emission peak at ~330 nm, which resembles the tryptophan buried in the hydrophobic core in the native protein and in agreement with previous observation by Elshereef at al.[10] This conclusion reinforces the stronger relevance of LV1 to the monomeric-IgG concentration. The LV2 encompasses regions of spectra attributed to tryptophan in the native (the peak at lower wavelengths) and in the non-native (peak at higher wavelengths) protein structure[14] with positive and negative correlations, respectively (Figure 5f). When tryptophan is further buried within the aggregates, the peak pertinent to non-native tryptophan increases. Thus, a reduction is observed in the scores of LV2. On the other hand, when aggregates are large enough to precipitate or are separated through centrifugation, the peak pertinent to non-native tryptophan decreases and consequently an increase is observed in the scores of LV2. However, the magnitude of

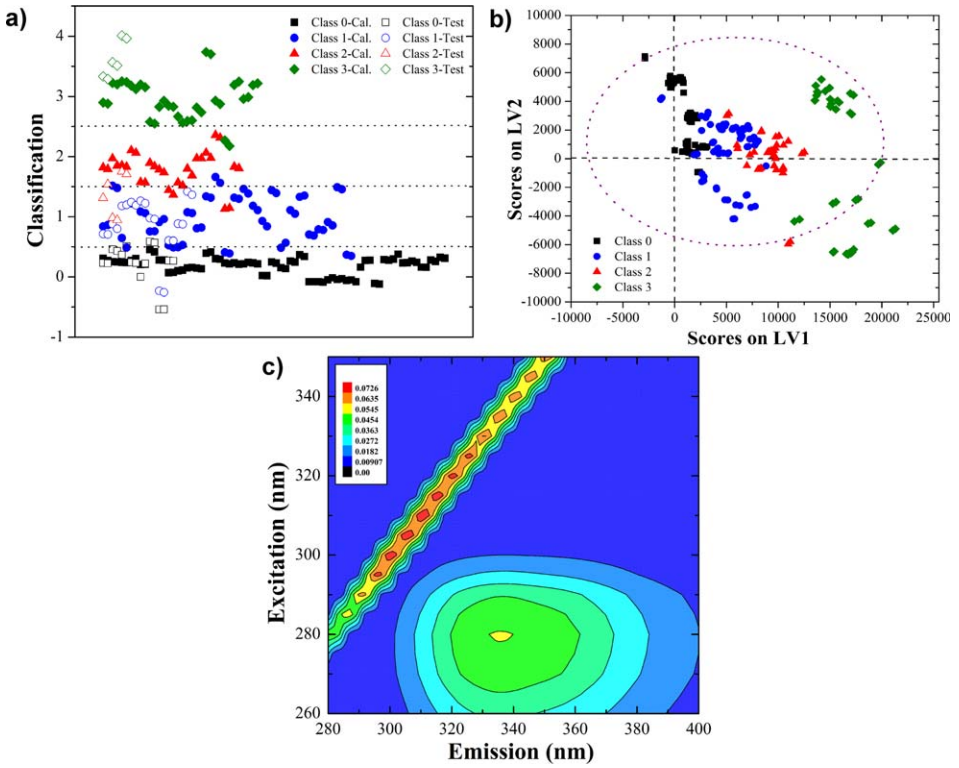

Figure 6. Results of the soft sensor for qualitative analysis: (a) Model classification of different samples (initial classes specified by symbols and colors); (b) Scores of LV2 versus LV1; (c) Contour plot of first loading.

scores of LV2 depends on the amount of monomer and type of aggregates in the sample and cannot be directly correlated to the aggregation percentage.

Soft sensor for quality control

The presence of oligomeric structures of Mab substantially diminishes the quality of the final product and can have profound negative impacts in the patient.[4,5,7,8] It imposes a need for a noninvasive approach for rapid quality control of final product. Fluorescence spectra, in conjunction with PLSR, were implemented to develop an empirical model capable of *at situ* classification of the final product into classes with different degrees of aggregation. To calibrate the model, samples were grouped into four classes that were defined based on the ratio of the percentage of the area under the SEC chromatograms' peak pertinent to tri/tetramers and that of the monomer IgG. For the purpose of model calibration, classes were assigned corresponding integer values: 0, 1, 2, and 3. A possible description of the classes corresponding to each one of these integer values could be as follow: (1) no to very small aggregates-acceptable sample; (2) mostly monomer with few aggregates-acceptable sample with caution; (3) equal amount of aggregates and monomers-require further evaluation; and (4) over half of the sample are aggregates-rejected sample. Clearly, the description of the classes is subjective and it could be modified if additional information such as the therapeutic efficacy of the Mab were available from other sources. The integer values describing the classes were used to form the response matrix Y that was regressed with respect to the input matrix containing the measured fluorescence spectra of the samples. As mentioned before, the scattering region of the spectra is attributed to larger particles and can provide information about the presence of oligomeric structures. Thus, for the formulation of this soft sensor the complete fluorescence spectra, encompassing both the scattering regions and region attributed to tryptophan, were implemented as the input matrix (X). To validate the model precision, randomly selected experiments were excluded from the calibration step. The cross validation procedure revealed that 6 LVs minimizes the RMSE-CV and generates a model capable of capturing 99.99% of input data set and explains 92.12% of output data with R^2_{CV} being equal to 90.1%. The subsequent model is capable of predicting the test set with the R^2_P of ~79%. Figure 6a illustrates the model predictions for the calibration set and validation set distinguished based on their specified class. Since the predictions of the model are continuous variables, i.e., continuous values between 0 and 3, thresholds were selected as boundaries of classes in order to assign to each predicted value one of the integer values between 0 and 3. It can be concluded from the Figure 6a that the model properly discriminates between classes with few samples being misclassified. Table 3 summarizes the confusion table for the calibration and validation sets where this table indicated the percentages of correct and incorrect detections obtained for each data set under study. From this table, the maximum probability for a true positive prediction (sensitivity) is observed for classes 0 and 3 for both calibration and validation sets, which supports the fact that the definite acceptance or rejection of the final product can be done with a high degree of certainty. The confusion table also indicates that the probability of true negative (selectivity) of different classes for calibration and validation sets are over 90%, except for class 1 for the validation set which is roughly 80%. Despite the presence of some misclassified samples, the model demonstrates high selectivity

Table 3. Confusion Table for Calibration and Test Set for the Soft Sensor for Quality Control

	Calibration Set				Test Set			
	Class 0	Class 1	Class 2	Class 3	Class 0	Class 1	Class 2	Class 3
Predicted as Class 0	74	8	0	0	13	2	0	0
Predicted as Class 1	0	43	4	0	3	18	3	0
Predicted as Class 2	0	3	26	2	0	0	3	0
Predicted as Class 3	0	0	0	32	0	0	0	6

and sensitivity which strengthens the potential of this model for quality control. The first LV plays the most crucial role accounting for ~84% of variation in X and predicting 90% of the response Y. From Figure 6b, the scores of LV2 versus LV1, it is observed that there is a strong correlation between the LV1 scores and classes. The physical relevance of LV1 is demonstrated by the contour plot of its loading, Figure 6c, which is dominated by the regions of spectra pertinent to scattering and, to a lesser degree, tryptophan-related emission.

Conclusion

On the basis of a comparison of HP-SEC and multi-wavelength fluorescence data it is demonstrated that intrinsic multi-wavelength fluorescence spectroscopy can be utilized for *at situ* quality control and quantitative monitoring of protein aggregation for different process conditions that typically occur during the downstream purification of monoclonal antibodies. The outlet of a bioreactor feed to downstream processing units typically consists of a complex mixture of a variety of proteins and nutrients that may affect the fluorescence signal and may complicate soft sensor development. In view of the complexity that may arise for samples taken directly from a bioreactor, as a preliminary evaluation of multi-wavelength fluorescence as a tool for *at situ* monitoring, the current study only focuses on single protein samples aggregated by various stress factors. As such, soft-sensors would have to be developed on an *ad hoc* basis for each particular process. To simulate the diverse processing conditions during downstream processing, IgG samples were stressed under various conditions that included concentration, pH, and temperature that resulted in different aggregation patterns. The impact of these factors on the fluorescence signal was investigated and compared with the results obtained from HP-SEC-based sizing. To systematically comprehend the influence of these operating factors individually or in conjugation with each other, on the degree of aggregation, a nonsymmetrical CCD was performed. A surface response model was fitted to the data, which helped to identify the operating regions where the propensity to aggregation was lower. An empirical predictive model was then established providing fast and accurate predictions of the monomer-IgG concentration in the sample. Finally, a model was developed using PLSR that successfully discriminated the samples into different classes corresponding to different degrees of aggregation. In view of the adverse impact of aggregation on therapeutic efficacy of antibodies, such a PLSR based model could be extremely beneficial for future online quality control. The feasibility of a fluorescence-based soft sensor for predicting both monomer concentration and aggregation patterns for a diverse set of operating conditions was demonstrated. The proposed approach offers potential for *at situ* monitoring of downstream processes for monoclonal antibody production. Future work will consider the applicability of the proposed soft sensor to samples involving mixtures of different proteins rather than single protein samples, as considered in the current work.

Acknowledgments

The authors would like to thank Natural Science and Engineering Research Council (NSERC) and Monoclonal Antibody Network (MabNet) for the funding of this research.

Literature Cited

1. Butler M, Meneses-Acosta A. Recent advances in technology supporting biopharmaceutical production from mammalian cells. *Appl Microbiol Biot*. 2012;96:885–894.
2. Sahin E, Grillo AO, Perkins MD, Roberts CJ. Comparative effects of pH and ionic strength on protein-protein interactions, unfolding, and aggregation for IgG1 antibodies. *J Pharm Sci-Us*. 2010;99:4830–4848.
3. Sadavarte RH, Ghosh R. A thermal-cycling method for disaggregating monoclonal antibody oligomers. *J Pharm Sci-Us*. 2014;103:870–878.
4. Mahler HC, Friess W, Grauschopf U, Kiese S. Protein aggregation: pathways, induction factors and analysis. *J Pharm Sci-Us*. 2009;98:2909–2934.
5. Printz M, Friess W. Simultaneous detection and analysis of protein aggregation and protein unfolding by size exclusion chromatography with post column addition of the fluorescent dye BisANS. *J Pharm Sci-Us*. 2012;101:826–837.
6. Wang W. Protein aggregation and its inhibition in biopharmaceutics. *Int J Pharm*. 2005;289:1–30.
7. Hawe A, Friess W, Sutter M, Jiskoot W. Online fluorescent dye detection method for the characterization of immunoglobulin G aggregation by size exclusion chromatography and asymmetrical flow field flow fractionation. *Anal Biochem*. 2008;378:115–122.
8. He F, Phan DH, Hogan S, Bailey R, Becker GW, Narhi LO, Razinkov VI. Detection of IgG aggregation by a high throughput method based on extrinsic fluorescence. *J Pharm Sci-Us*. 2010;99:2598–2608.
9. Abbas SA, Gaspar G, Sharma VK, Patapoff TW, Kalonia DS. Application of second-derivative fluorescence spectroscopy to monitor subtle changes in a monoclonal antibody structure. *J Pharm Sci-Us*. 2013;102:52–61.
10. Elshereef R, Budman H, Moresoli C, Legge RL. Fluorescence spectroscopy as a tool for monitoring solubility and aggregation behavior of beta-lactoglobulin after heat treatment. *Biotechnol Bioeng*. 2006;95:863–874.
11. Ohadi K, Aghamohseni H, Legge RL, Budman HM. Fluorescence-based soft sensor for at situ monitoring of chinese hamster ovary cell cultures. *Biotechnol Bioeng*. 2014;111:1577–1586.
12. Kumar V, Sharma VK, Kalonia DS. Second derivative tryptophan fluorescence spectroscopy as a tool to characterize partially unfolded intermediates of proteins. *Int J Pharm*. 2005;294:193–199.
13. Poole RA, Hawe A, Jiskoot W, Braeckmans K. Fluorescence spectroscopy to characterize protein aggregates and particles. Analysis of Aggregates and Particles in Protein Pharmaceuticals. NJ, USA: Wiley; 2012:201–226.
14. Elshereef R, Budman H, Moresoli C, Legge RL. Fluorescence-based soft-sensor for monitoring beta-lactoglobulin and alpha-lactalbumin solubility during thermal aggregation. *Biotechnol Bioeng*. 2008;99:567–577.

15. Herbert S, Riou, NM, Devaux MF, Riaublanc A, Bouchet B, Gallant DJ, Dufour É. Monitoring the identity and the structure of soft cheeses by fluorescence spectroscopy. *Lait*. 2000;80:621–634.
16. Rathore AS, Bhambure R, Ghare V. Process analytical technology (PAT) for biopharmaceutical products. *Anal Bioanal Chem*. 2010;398:137–154.
17. Teixeira AP, Oliveira R, Alves PM, Carrondo MJT. Advances in on-line monitoring and control of mammalian cell cultures: supporting the PAT initiative. *Biotechnol Adv*. 2009;27:726–732.
18. Wold S, Ruhe A, Wold H, Dunn WJ. The collinearity problem in linear-regression—the partial least-squares (Pls) approach to generalized inverses. *Siam J Sci Stat Comp*. 1984;5:735–743.
19. Hatti-Kaul R, Mattiasson B. Isolation and Purification of Proteins. NY, USA: CRC Press; 2003.
20. Geladi P, Kowalski BR. Partial least-squares regression—a tutorial. *Anal Chim Acta*. 1986;185:1–17.
21. Liu CD, Vailhe C, Samuel N, Min Q. A simple and reproducible approach to characterize protein stability using rheology. *Int J Pharm*. 2009;374:1–4.

Manuscript received Jan. 23, 2015, and revision received May 29, 2015.